COMRADES determines in vivo RNA structures and interactions

Determining *in vivo* RNA structures with conformation capture using clickable psoralen

Omer Ziv[1,*,#] , Marta M. Gabryelska[2,*], Aaron T.L. Lun[3], Luca F.R. Gebert[4], Jessica Sheu-Gruttadauria[4], Luke W. Meredith[5], Chun Kit Kwok[6], Ian J. Macrae[4], Ian Goodfellow[5], John C. Marioni[3], Grzegorz Kudla[2,#] and Eric A. Miska[1,7,#]

[1]Wellcome Trust/Cancer Research UK Gurdon Institute and Department of Genetics, University of Cambridge, Cambridge, United Kingdom; [2]MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, Scotland, UK; [3]Cancer Research UK Cambridge Institute, University of Cambridge, Li Ka Shing Centre, Cambridge, United Kingdom; [4]Department of Integrative Structural and Computational Biology, The Scripps Research Institute, La Jolla, CA, USA; [5]Division of Virology, Department of Pathology, University of Cambridge, Addenbrooke's Hospital, Cambridge, UK; [6]State Key Laboratory of Pathogen and Biosecurity, Beijing Institute of Microbiology and Epidemiology, Academy of Military Medical Sciences, Beijing, China; [7]Department of Chemistry, City University of Hong Kong, Kowloon Tong, Hong Kong SAR, China; [8]Wellcome Sanger Institute, Hinxton, Cambridge, UK.

* These authors contributed equally to this work.

[#] Corresponding authors: E.A.M. (eric.miska@gurdon.cam.ac.uk), G.K. (gkudla@gmail.com), O.Z. (omer.ziv@gurdon.cam.ac.uk).

RNA structural flexibility underlies fundamental biological processes, including splicing, translation, and the lifecycle of RNA viruses. However, there are no experimental or computational methods to explore the multiple conformations adopted by RNAs *in vivo*. We developed Crosslinking Of Matched RNAs And Deep Sequencing (COMRADES) for in-depth conformation capture of RNA and retrieval of RNA structural ensembles. Using COMRADES, we determined the architecture of the Zika virus RNA genome inside cells, and revealed multiple interactions with human non-coding RNAs.

Structure probing techniques have been proven valuable for studying cellular and viral RNA structure[1–9]. Inside cells however, RNA is engaged in multiple processes including splicing, translation, and regulation by RNAi, suggesting high structural plasticity. Recent methods have utilized proximity ligation to reveal RNA base-pairing within cells [10–16]. Nevertheless, the ability to assess the structural plasticity of RNA and report the ensemble of coexisting conformations has not been reached, due to insufficient probing depth and lack of appropriate computational algorithms. We developed a method—Crosslinking Of Matched RNAs And Deep Sequencing (COMRADES)—that couples an improved *in vivo* probing of RNA base-pairing with selective RNA capturing. We additionally established an algorithm for assessing the structural complexity of RNA inside cells (Fig 1a).

COMRADES utilizes a cell-permeable, azide modified, psoralen derivative (Psoralen-triethylene glycol azide) to facilitate coupling of two effective affinity capturing steps, while overcoming the limited cell permeability of biotin labelled psoralen (Supplementary Fig. 1a). The azide group does not affect the psoralen

crosslinking properties (Supplementary Fig. 1a). Following *in vivo* crosslinking, an RNA of interest is selectively captured, reaching enrichment level of nearly 1000-fold (Fig. 1b, Supplementary Fig. 1b-c). The RNA is then fragmented, and a copper free click-chemistry reaction links a biotin moiety to *in vivo* crosslinked regions, enabling a second streptavidin-based affinity selection of crosslinked regions (Fig. 1c, Supplementary Fig. 1d). The resulting RNA is split: One half is proximity ligated to create RNA chimeras, following reversal of the crosslink to enable high-throughput sequencing and assessment of the base-pairing (Fig. 1d, Supplementary Fig. 1e). The other half is used as a control, in which reversal of the crosslink precedes the proximity ligation. COMRADES and control samples contain essentially identical RNA composition, ensuring accurate assessment of artificial chimeric reads originated from random ligation or reverse transcription errors. COMRADES's dual enrichment substantially increases structure probing depth of selected RNA, thus enabling an unbiased and global view of coexisting conformations. COMRADES yields high levels of ligated chimeric reads (8%), whereas these are kept 4 fold lower in the controls, or in a non-crosslinked samples (Fig. 1e). We successfully reported on the known ribosomal RNA structure with high sensitivity (Supplementary Fig. 2), while spurious interactions between cytoplasmic and mitochondrial ribosomal RNA subunits occurred at a very low level (Fig. 1f). The robustness of COMRADES is further demonstrated by its high reproducibility (Fig. 1g-h).

RNA viruses utilize RNA base-pairing to regulate various aspects of their life cycle[17–21]. Inside the host cell however, the full-length architecture of RNA genomes and their interactions with the host transcriptome are largely unknown. We used COMRADES to determine RNA base-pairing along the 10.8 kilobases-long single-

stranded RNA genome of Zika virus (ZIKV) from the *Flavivirus* genus inside human cells. We identified 1.7 million non-redundant chimeric reads corresponding to the structure of the ZIKV genome (Fig. 1g-h, Supplementary Fig. 3a). This high probing coverage is valuable for analysing multiple coexisting conformations. Previous work mainly identified RNA structures in the untranslated regions (UTRs) of flaviviruses, while leaving 95% of the genome unexplored[17–21]. COMRADES identified base-pairing along the entire genome and between the open reading frame (ORF) and the UTRs (Supplementary Fig. 3b, 4a). Nearly 80% of the identified interactions span a distance of less than 1,000 nucleotides (nt), implying local structure with a certain degree of three-dimensional compaction (Supplementary Fig. 3c). Both short- and long-range interactions were supported by reproducible, well-defined clusters of chimeric reads, ligated in 5'-3' and 3'-5' orientations (Supplementary Fig. 3a), and showed strong evidence of base-pairing when analyzed with the hybrid-min RNA-folding algorithm as compared to a shuffled-chimeras control (Wilcoxon test p-value < 0.0001). COMRADES therefore enables a deep and comprehensive analysis of RNA base-pairing.

During replication, the genome of flaviviruses undergoes a global conformational change mediated by a long-distance base-pairing between the 5' and 3' cyclization sequences[17,18] (5' CS and 3' CS respectively), as well as the upstream and downstream of AUG regions (UAR and DAR respectively)[19,20]. COMRADES detected extensive and highly specific base-pairing between the known cyclization elements, therefore demonstrating genome cyclization inside cells (Fig. 2a-c, Supplementary Fig. 4b). COMRADES further refined the nature of the base-pairing associated with genome cyclization, by identifying contact regions upstream of the 5'

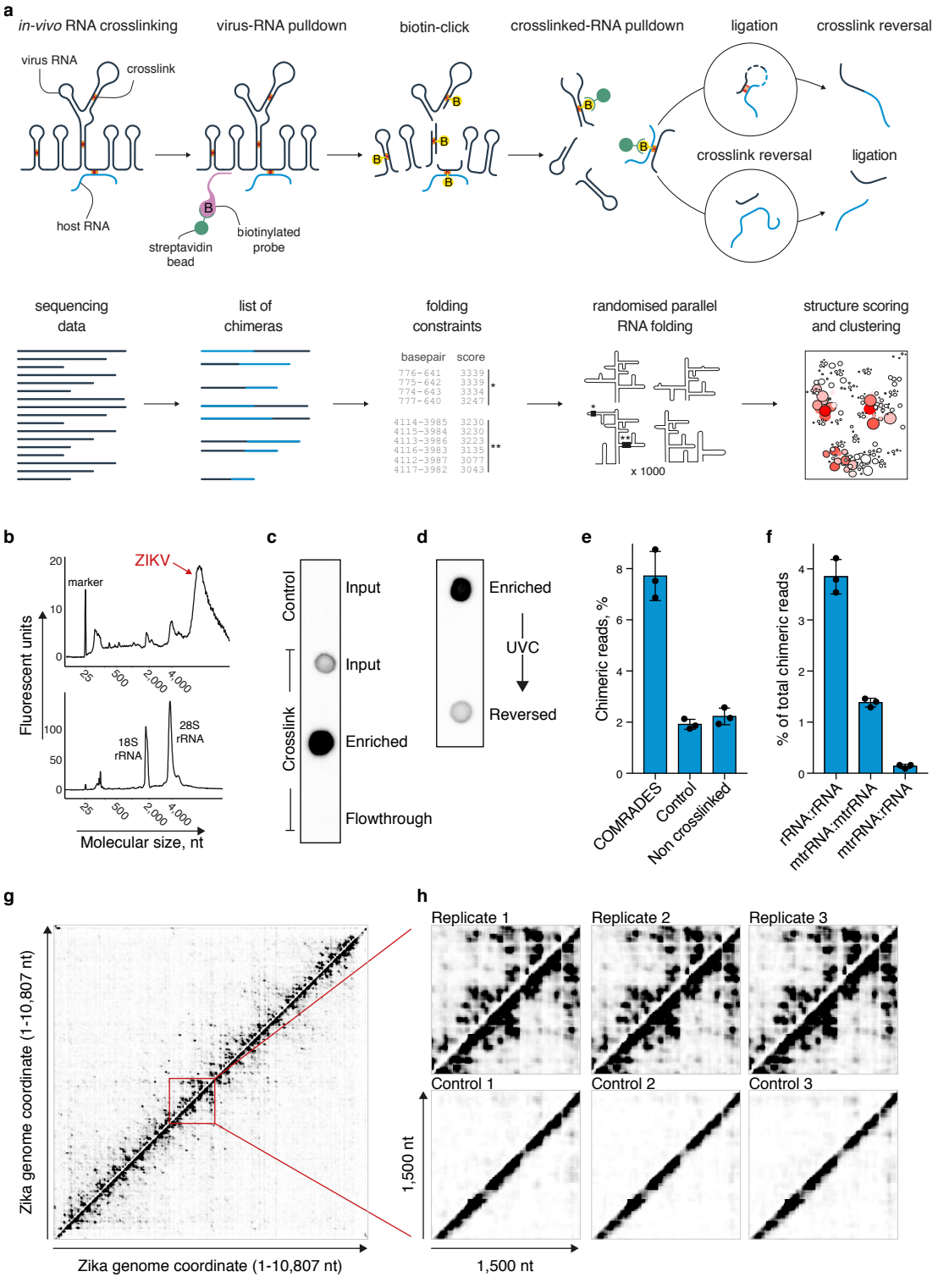
UAR and downstream of the 3' UAR (Fig. 2a,c). We additionally detected an alternative 5' UTR conformation where stem-loops A and B (SLA and SLB respectively) are not formed but rather engaged in long-distance base-pairing with the downstream envelope coding sequence (Fig. 2d-f). COMRADES confirmed the existence of previously defined functional RNA pseudoknots including the dumbbell (DB) pseudoknot[22], the downstream of 5' cyclization sequence-pseudoknot (DCS-PK)[23], and the SL1 pseudoknot[24] (Supplementary Fig 4c). Overall, COMRADES identified nearly all previously known flavivirus RNA structures and has further defined critical base-pairing involving the UTRs.

Our intra-viral RNA-RNA interaction map revealed the presence of multiple mutually exclusive RNA structures, where one region alternately base-pairs with several other regions. The averaged Shannon entropy per nucleotide was 5.9 bits, implying high folding plasticity (Supplementary Fig. 5a). We found a strong inverse correlation between the degree of experimental support for base paired regions and their entropy (Supplementary Fig. 4a, 5b-d). To explore the ensemble of alternative structures, we developed an algorithm to computationally fold ~1,000 nucleotide-long regions using randomly selected subsets of high-confidence mutually compatible folding constraints derived from the *in vivo* data. For each region, a set of 1,000 structures was generated. The validity of this approach is demonstrated by the clear correlation between the thermodynamic stability and the number of reads supporting each structure (Fig 2g, Supplementary Fig. 6). Nevertheless, the most thermodynamically favoured structures gained only moderate experimental support, implying the additional impact of the cellular environment on RNA folding[3,4,16,25,26].

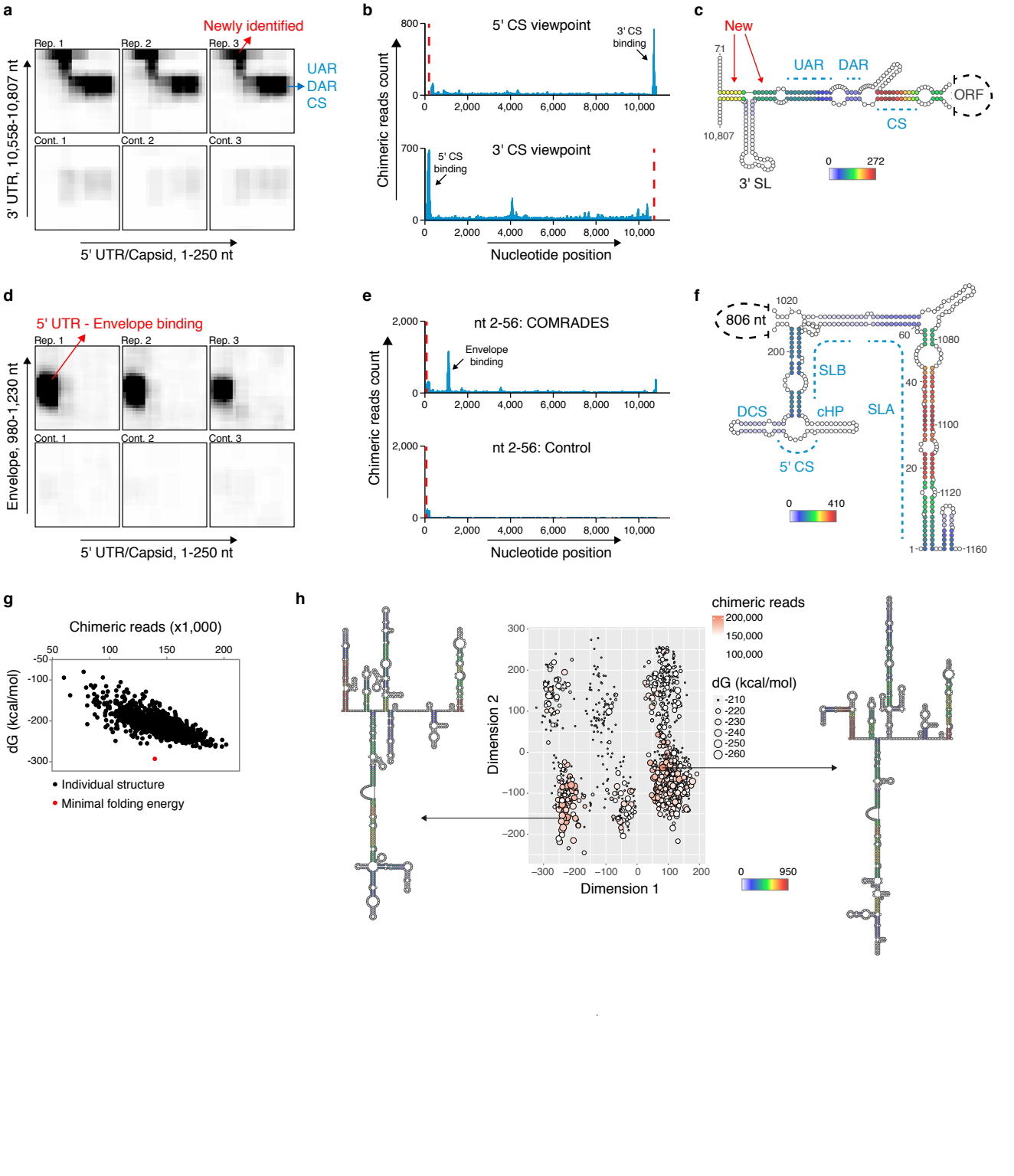
126 We further computed the degree of similarity between all pairs of structures
127 and applied multidimensional scaling to cluster structures based on their similarity.
128 The presence of separated well defined clusters reflects the occurrence of
129 alternative conformations (Fig. 2h, Supplementary Fig. 7a). As a control, we
130 randomly shuffled the interacting RNA partners between the chimeric reads; the
131 resulting shuffled structures clustered separately from the structures recovered by
132 COMRADES (Supplementary fig. 8). A single ZIKV structure typically accounted for
133 ~30% of the *in vivo* observed interactions, whereas a reduced set of 5 structures
134 was sufficient to capture 80-90% of the *in vivo* data (Supplementary Fig. 7b). Our
135 analysis suggests that the intracellular folding complexity of the ZIKV genome might
136 be explained by postulating the coexistence of a small subset of alternative
137 conformations.
138
139 Viral RNAs have an inherent capacity to form specific interactions through
140 base-pairing with host RNAs[27], however little is known about the prevalence of such
141 interactions. COMRADES revealed multiple interactions between the ZIKV genome
142 and human small regulatory RNAs (Fig. 3a). We found site-specific interactions
143 between the ZIKV ORF and the U1 small nuclear RNA (snRNA, Fig. 3b), plausibly
144 affecting host splicing. We similarly detected highly site-specific interactions between
145 the ZIKV ORF and tRNAs Glu^{116} and Gly^{34} (Fig. 3c). We identified several
146 interactions between the ZIKV genome and human microRNAs, including miR-21,
147 miR-19, miR-512, miR-515, and miR-1323 (Fig. 3d). While previous psoralen-based
148 methods failed to effectively capture interactions involving miRNAs[10–12],
149 COMRADES successfully discovered miRNA base-pairing (Fig. 3a,d). COMRADES
150 indicated non-canonical base-pairing between the 5' CS of ZIKV and the seed region

of miR-21 (Fig. 3e, adjusted p-value $1.0E^{-13}$). The miR-21 interaction with the ZIKV genome was further proved significant using an independent analysis pipeline (FDR $3.0E^{-25}$, supplementary methods). *In vitro* synthesized miR-21 failed to bind the ZIKV 5' CS on its own, while pre-loading miR-21 onto purified Argonaute2 (Ago2) facilitated a strong and sequence specific interaction (Supplementary Fig. 9), supporting the involvement of Ago2 in this base-pairing. CRISPR/Cas9 deletion or antisense inhibition of miR-21 in human cells decreased the intracellular level of the ZIKV genome (Supplementary Fig. 10a-d) or the ZIKV envelope protein (Supplementary Fig. 10e-j). Abrogating the miR-21 binding ability of a ZIKV replicon through point mutations renders it insensitive to miR-21 antisense inhibition (Supplementary Fig. 10k), indicating that miR-21 acts through direct interaction with the 5' CS.

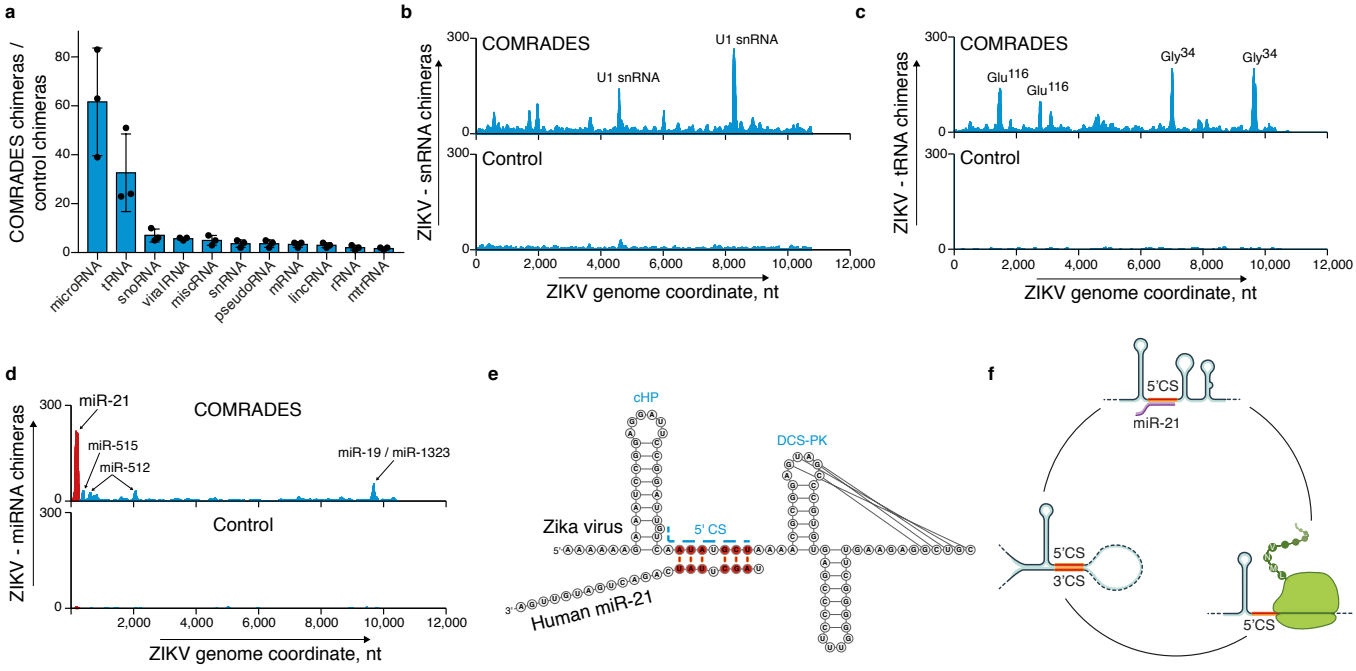
In summary, COMRADES revealed the highly dynamic nature of an RNA genome inside cells, and its ability to engage in base-pairing with multiple host regulatory RNAs. The involvement of the conserved 5' CS of ZIKV in genome cyclization, capsid translation and miR-21 binding further demonstrates the structural complexity of RNA genomes inside cells (Fig. 3f). The general applicability of COMRADES provides an opportunity to undertake a complete and unbiased analysis of the dynamic nature of RNA inside cells and can be utilized to investigate the structure and interacting partners of any cellular or foreign RNA in any species.



Ziv *et al.*, Fig. 1



Ziv *et al.*, Fig. 2



Ziv *et al.*, Fig. 3

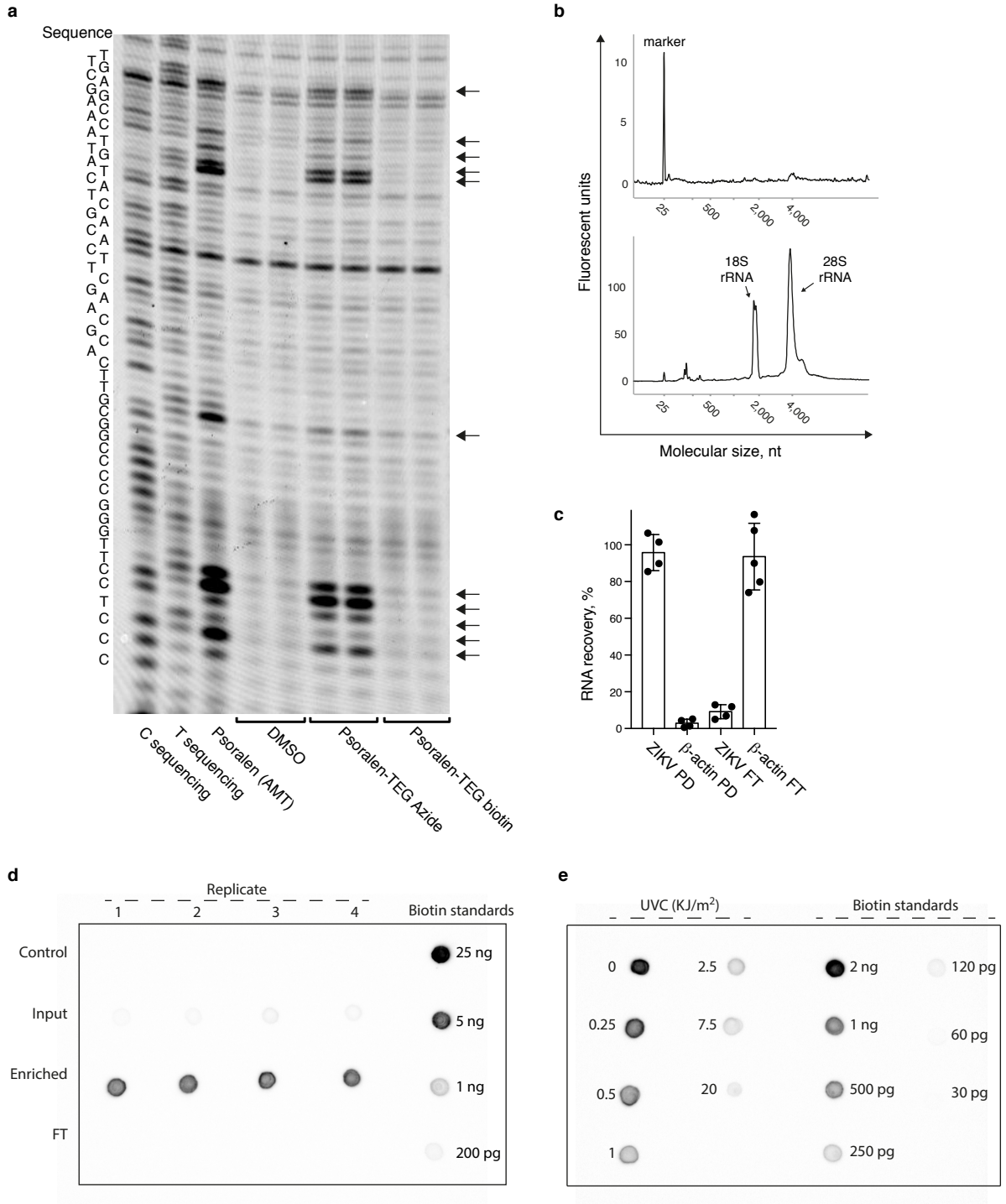
Figure legends

Fig. 1: **COMRADES methodology.** a, Experimental and computational workflows (upper and lower panels respectively). B: biotin; b, Virus-RNA pulldown, demonstrated by bioanalyzer profiles of ZIKV enriched RNA (upper panel) and input RNA (lower panel). c, Crosslinked-RNA pulldown, demonstrated by dot blot of biotin labelled RNA from COMRADES experiment. d, Crosslink reversal, demonstrated as described in (c). UVC: short-wavelength UV. e, % of chimeric reads in COMRADES and control libraries. Error bars represent standard deviation of 3 biological replicates. f, Probed interactions among cytoplasmic and mitochondrial ribosomal RNA subunits. rRNA: cytoplasmic ribosomal RNA; mtrRNA: mitochondrial ribosomal RNA. Error bars represent standard deviation of 3 biological replicates. g, Heat map of ZIKV RNA-RNA interactions. Each dot represents an interaction between the genomic coordinates on the x and y axes. Chimeras ligated in 5'-3' and 3'-5' orientations are plotted above and below the diagonal respectively. h, Zoom-in of a selected 1,500 nt region from (a).

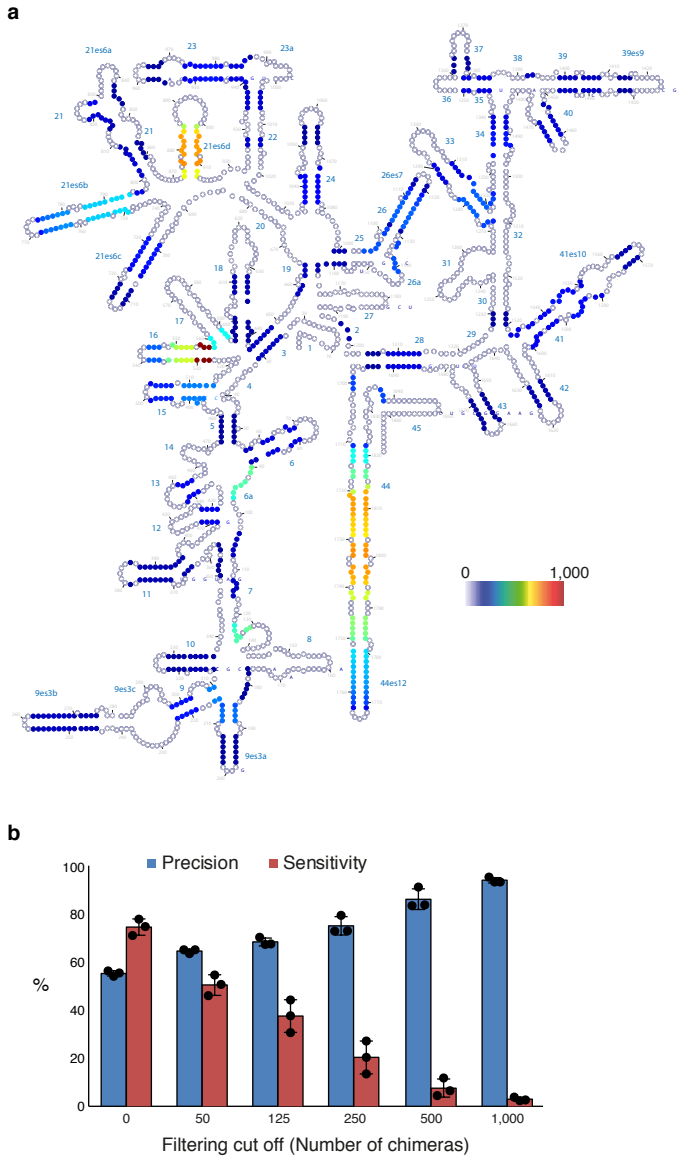
Fig. 2: **The genomic structure of ZIKV inside human cells.** a, Heatmap of RNA-RNA interactions between cyclization elements. Rep: biological replicate; cont: control. b, Viewpoint histograms showing binding positions of the cyclization sequences along the ZIKV genome. Viewpoint regions are marked by dashed red lines. c, Probed interactions along the circular genome conformation. Colours representing numbers of non-redundant chimeric reads supporting each interaction. New: newly identified base-pairing. d, Heatmap of RNA-RNA interactions between the 5' UTR and the envelope coding region. e, Viewpoint histogram showing binding of nucleotides at position 2-56 along the ZIKV genome. f, Newly identified 5' UTR

structure. Colour code as described in (c). cHP: capsid-coding region hairpin element; DCS: downstream of 5' cyclization sequence. g, Computationally predicted structures for nt 2288-3323. dG: folding energy. h, Clustering and prediction of alternative structures of nt 2288-3323. Colour code as described in (c).

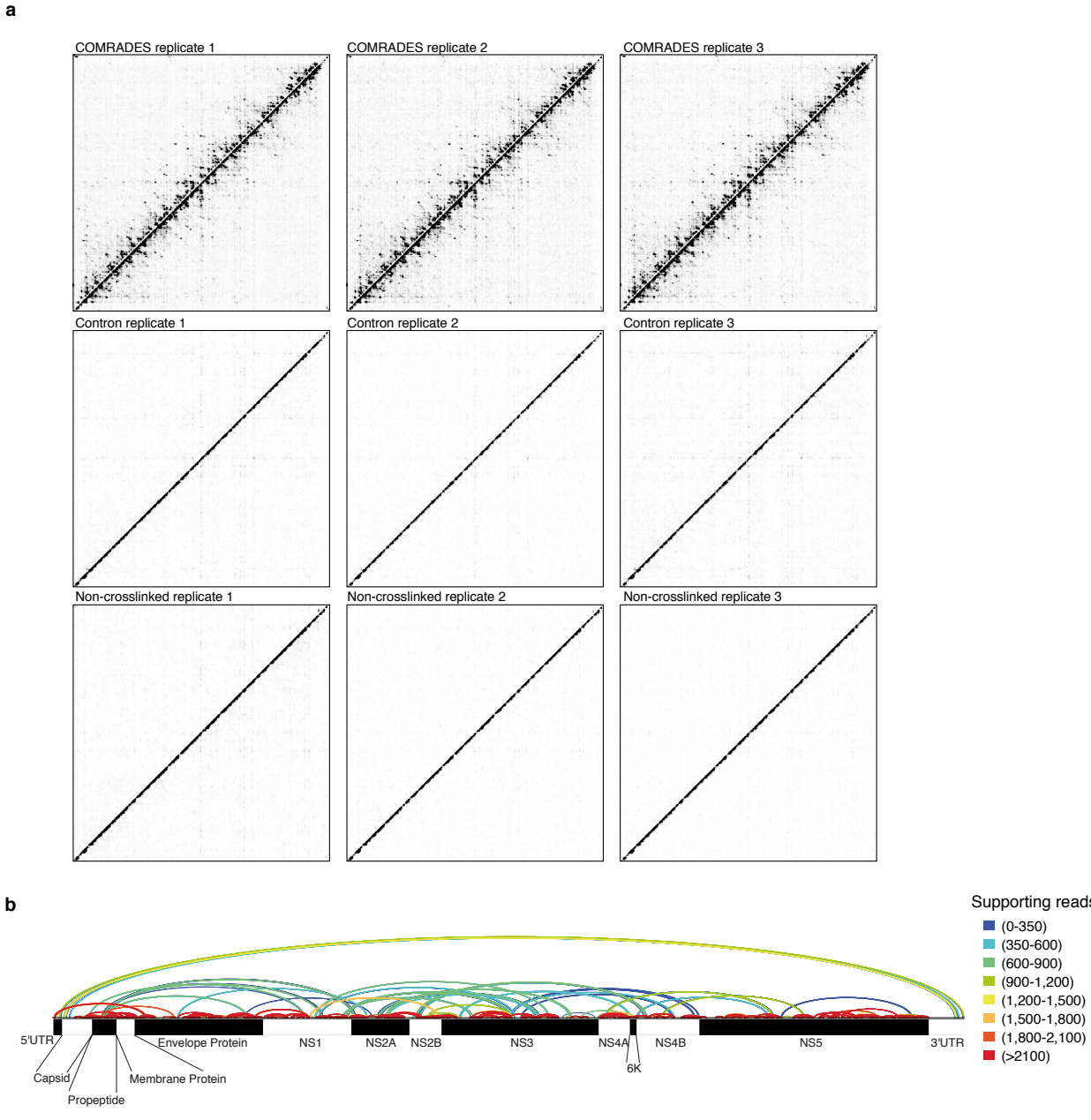
Fig. 3: **Host-virus RNA-RNA interactions.** a, Human RNA species interacting with the ZIKV genome. Error bars represent standard deviation of 3 biological replicates. b-d, Site specific base-pairing between the ZIKV genome and snRNAs (b), tRNAs (c), and specific miRNAs (d) in COMRADES libraries and controls. e, COMRADES determined base-pairing between ZIKV and miR-21. f, A model of the ZIKV 5' CS engaged in three separate functions. Ribosome and nascent polypeptide are marked in green.



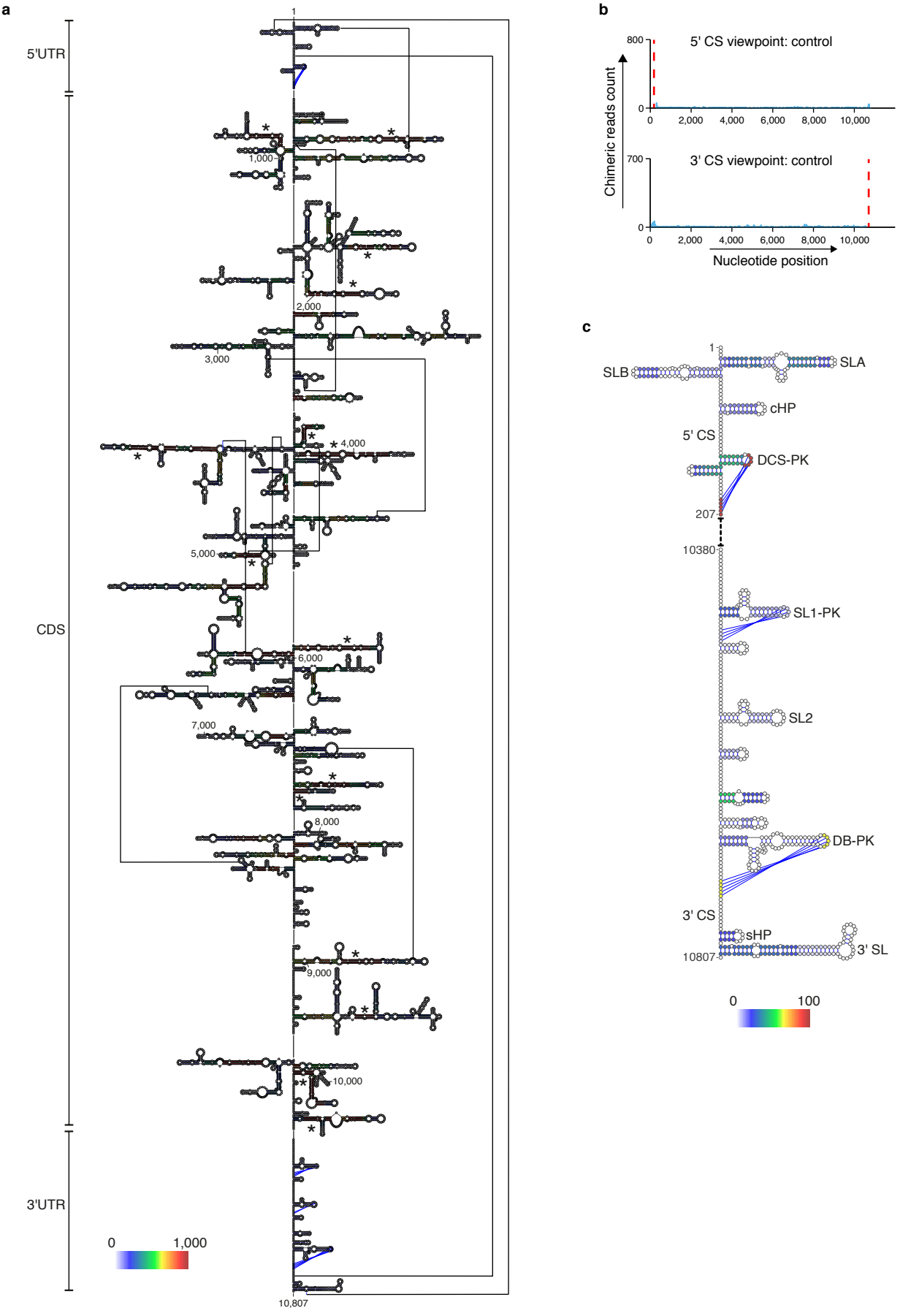
Ziv *et al.*, Supplementary Fig. 1



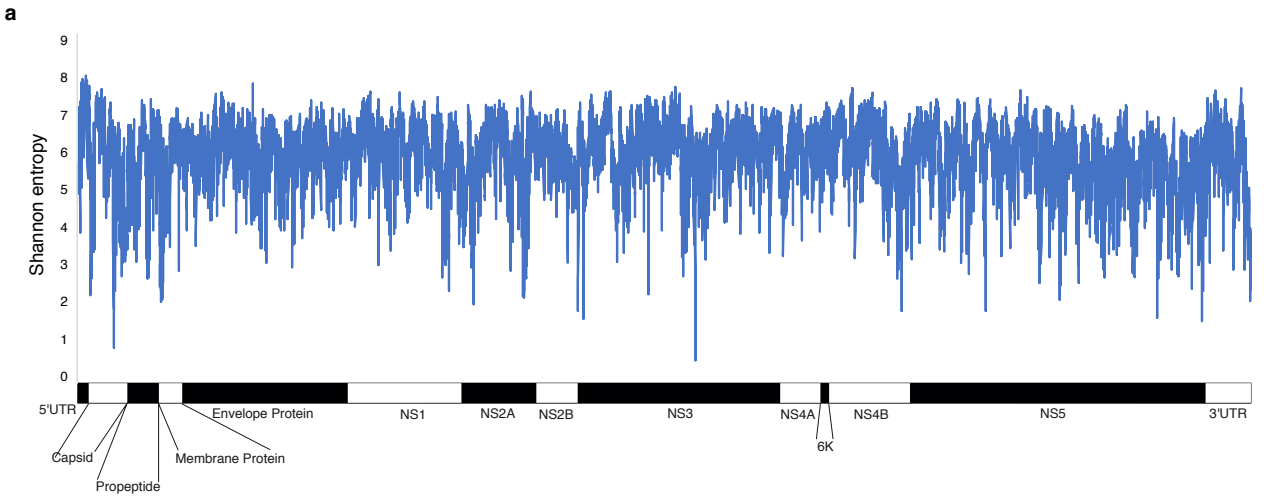
Ziv *et al.*, Supplementary Fig. 2

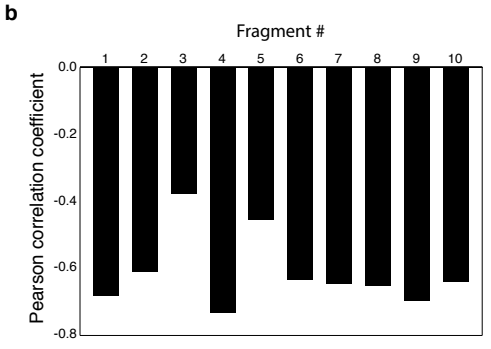


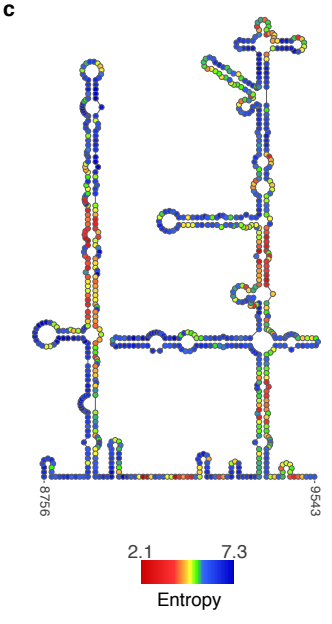
Ziv *et al.*, Supplementary Fig. 3

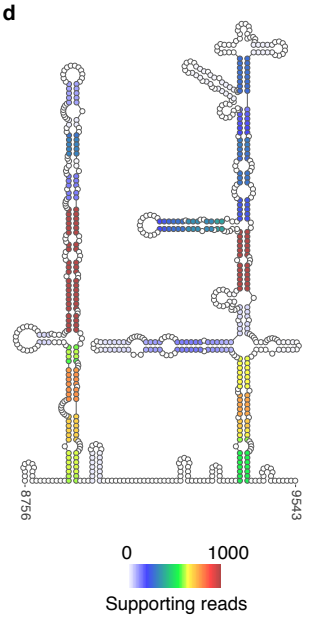


Ziv *et al.*, Supplementary Fig. 4

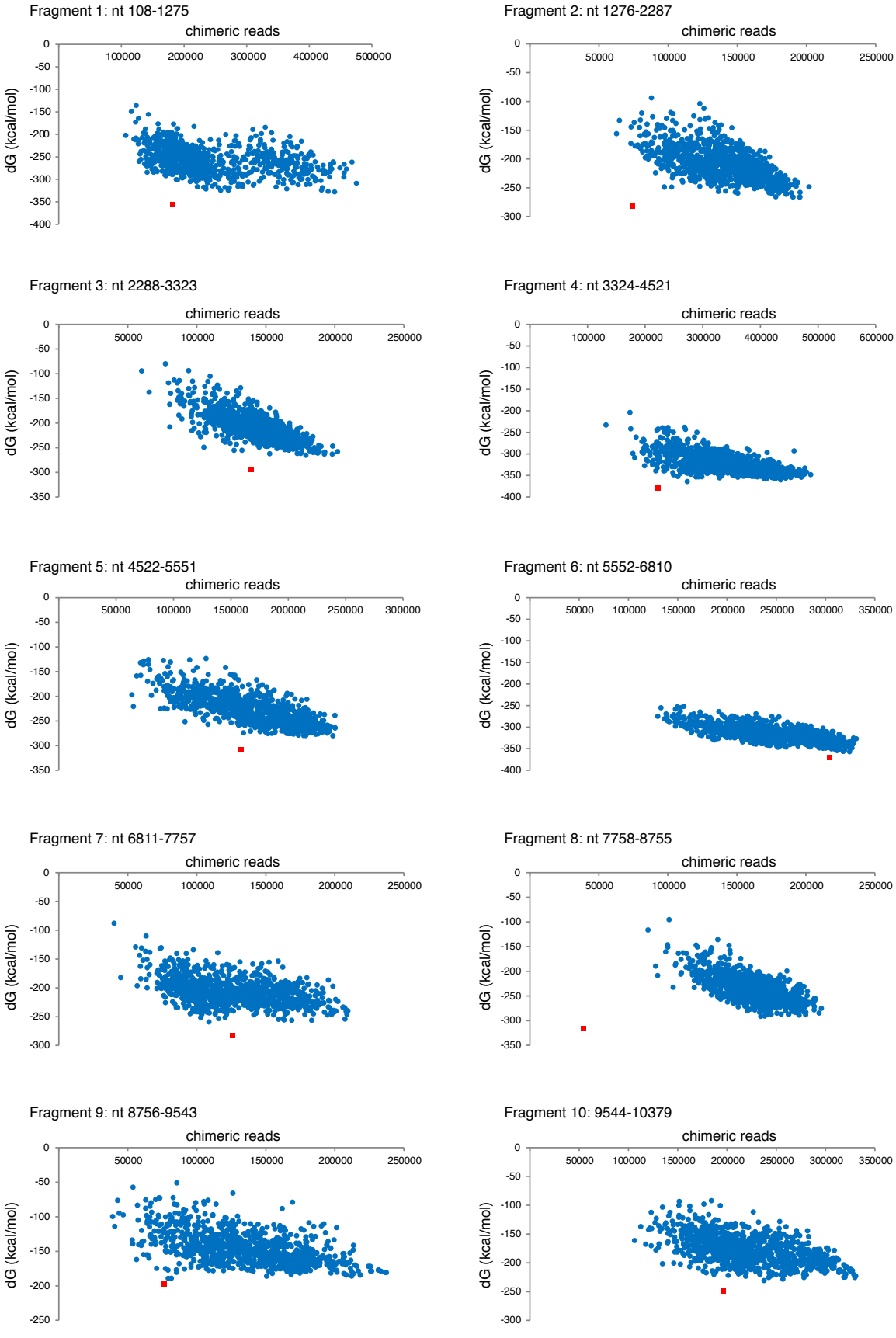




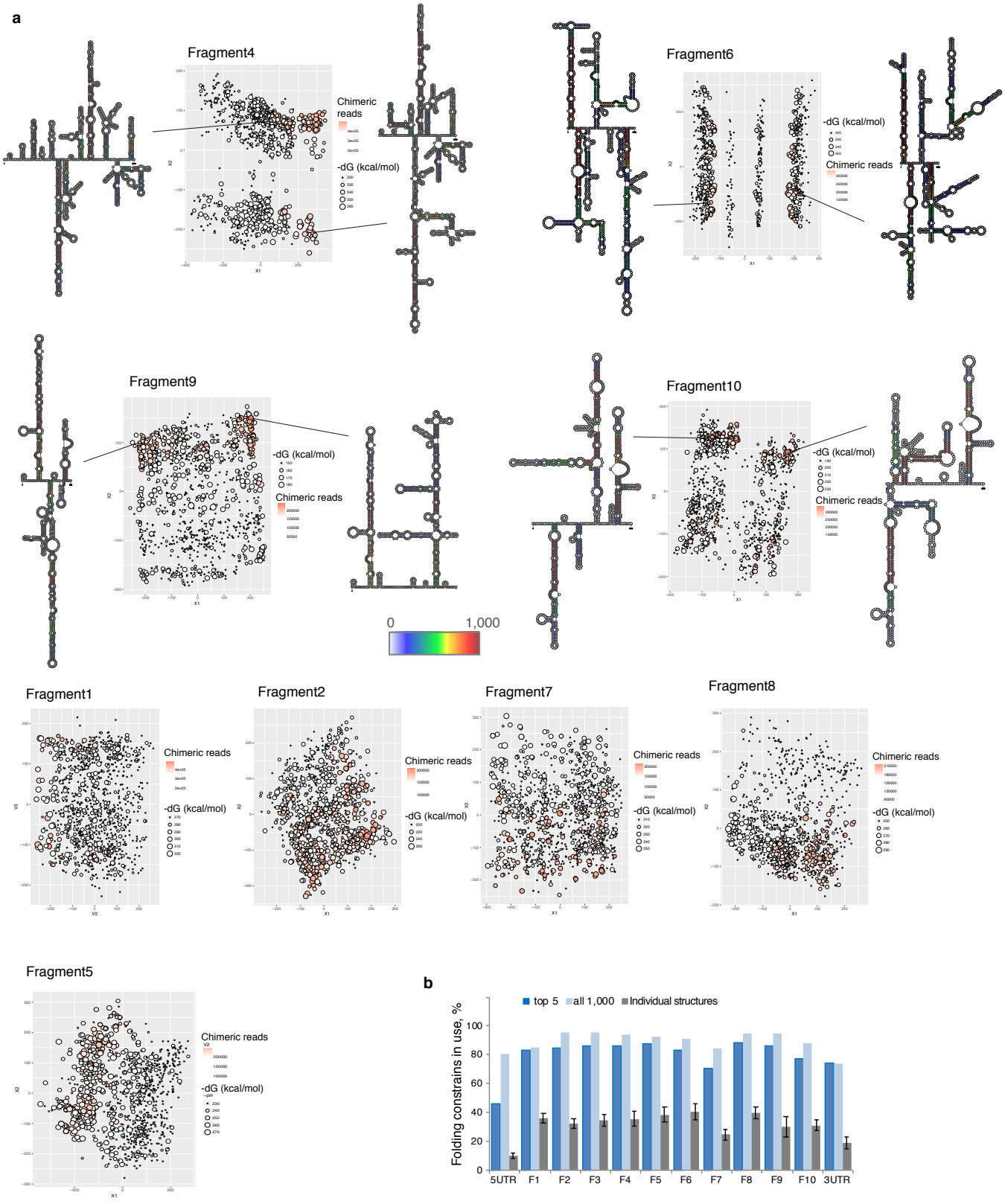




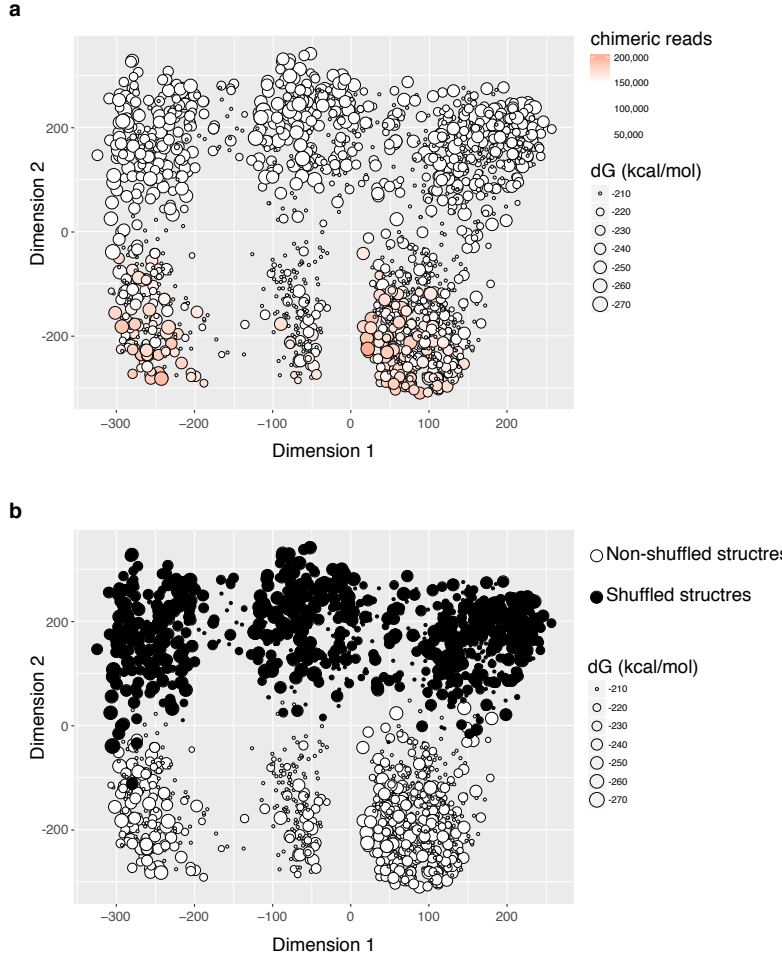
Ziv *et al.*, Supplementary Fig. 5



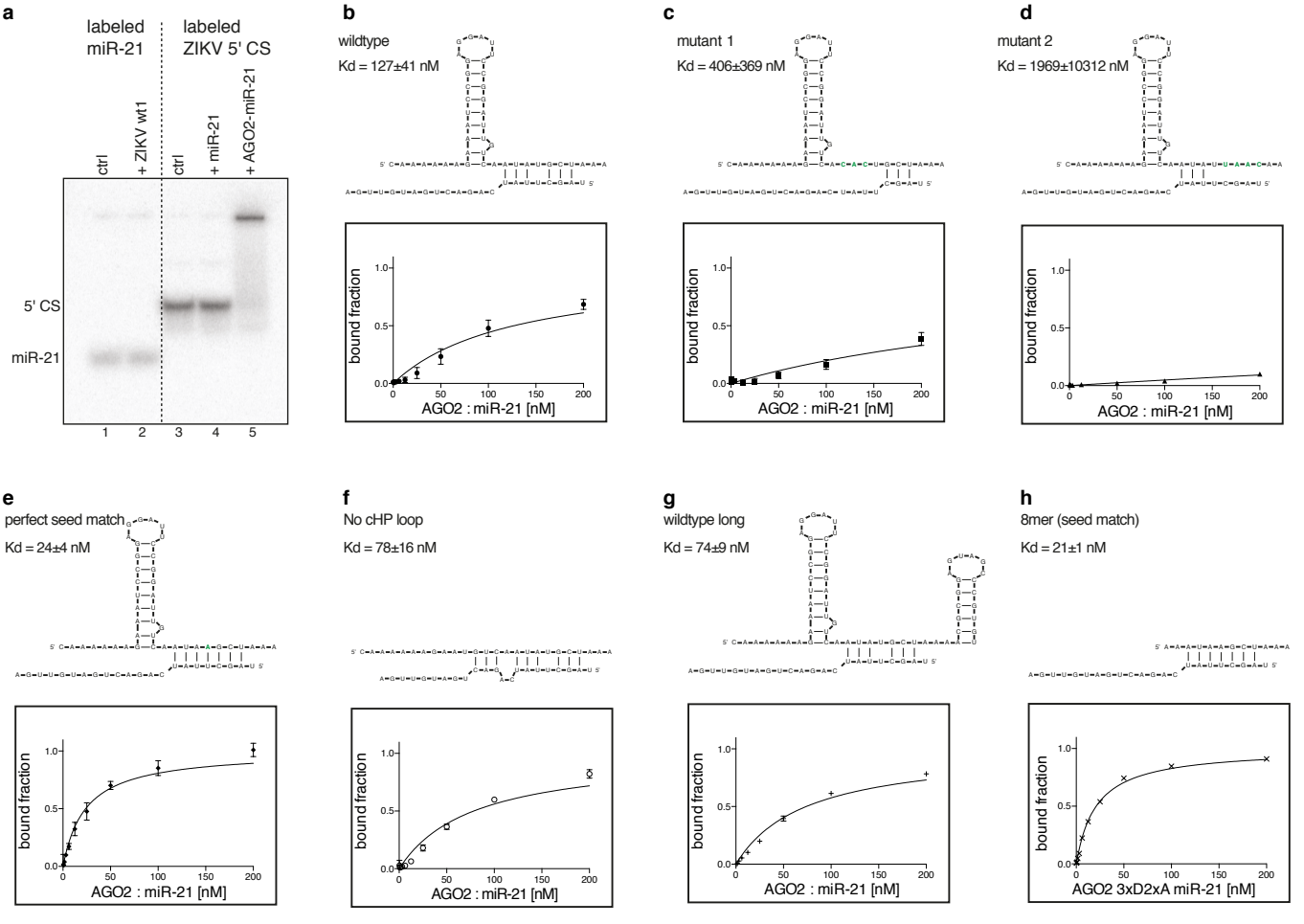
Ziv *et al.*, Supplementary Fig. 6



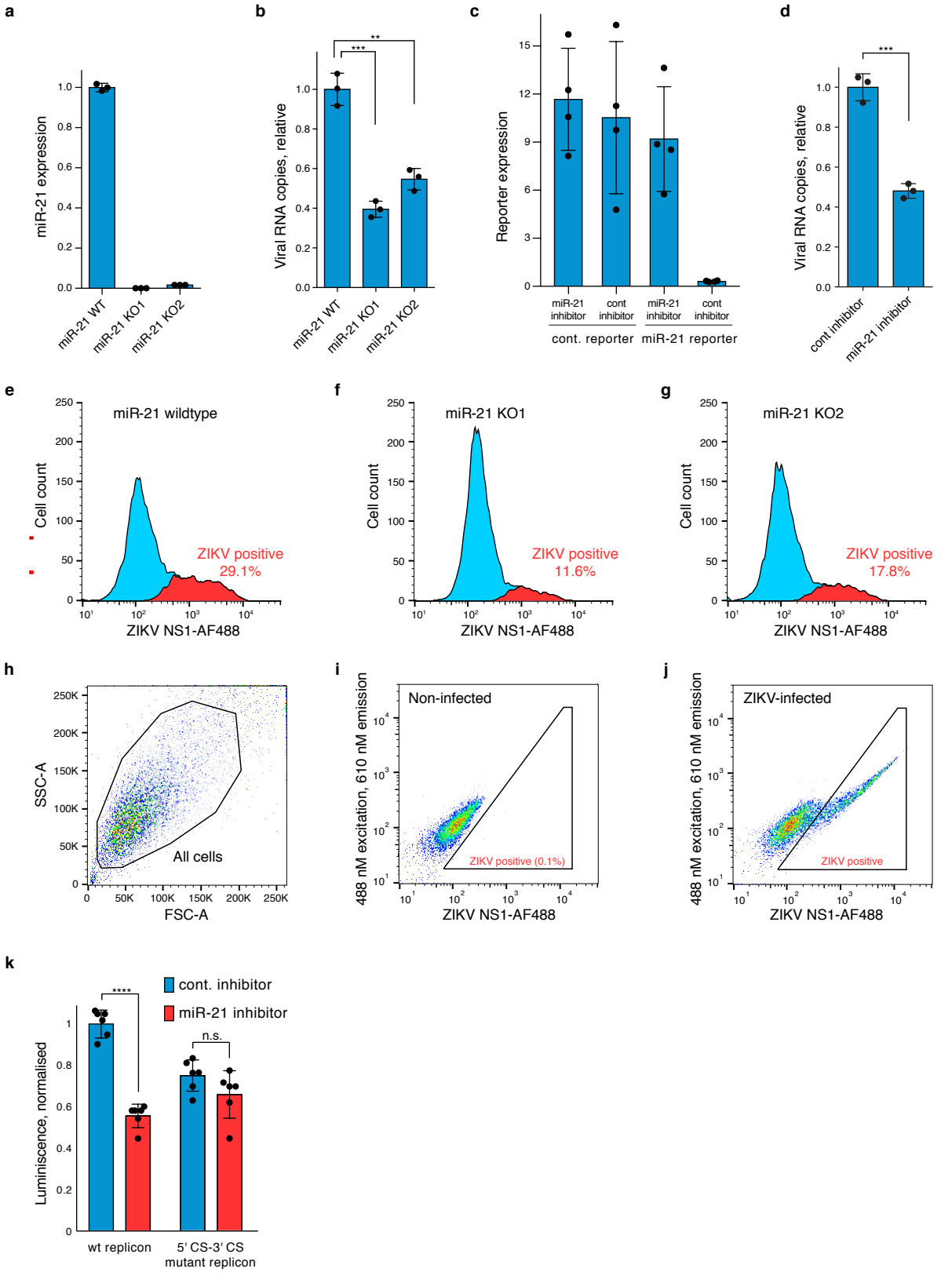
Ziv *et al.*, Supplementary Fig. 7



Ziv *et al.*, Supplementary Fig. 8



Ziv *et al.*, Supplementary Fig. 9



Ziv *et al.*, Supplementary Fig. 10

Supplementary figure legends

Supplementary Fig. 1: COMRADES assay development. a, Reverse transcription stalling assay indicating *in vivo* crosslinking positions in the 5.8S rRNA from cells treated with Psoralen (AMT), Psoralen-TEG azide, Psoralen-TEG biotin, or DMSO control. Arrows indicate stalling events. b, Bioanalyzer RNA profiles of ZIKV enriched (upper panel) or input (lower panel) RNA from non ZIKV-infected cells, as a control to Fig. 1b. c, Enrichment of ZIKV RNA measured by TaqMan PCR. Beta-actin serves as a control RNA. d-e, Non-cropped dot-blots corresponding to Fig. 1c-d. A shorter exposure is shown.

Supplementary Fig. 2: COMRADES validation. a, *In vivo* detected interactions overlaid on the Ribovision 18S phylogenetic ribosomal RNA secondary structure. Colour-code is indicative of the number of supporting chimeras for each base-pair. b, Precision and sensitivity of ribosomal RNA base-pairing detection by COMRADES. Analysis is based on ZIKV enriched libraries, therefore the sensitivity of COMRADES is expected to be underestimated by this analysis.

Supplementary Fig. 3: Intra-viral RNA interactions. a, Heat maps of ZIKV RNA-RNA interactions in crosslinked libraries and controls. Chimeras ligated in 5'-3' and 3'-5' orientations are plotted above and below the diagonal respectively. b, Arch plot representation of short and long-range RNA-RNA interactions along the ZIKV genome. Colours representing the number of non-redundant chimeric reads supporting each interaction. c, Distribution of RNA-RNA interactions by nucleotide distance between interacting chimeric partners. Left pie chart shows distribution of all

250 interactions; right pie chart shows distribution of interactions that span less than
251 1,000 nucleotides.
252
253 **Supplementary Fig. 4: The genomic structure of ZIKV inside human cells.** a,
254 The ZIKV genome conformation with the highest chimeric-reads support. Colour
255 code is indicative of the number of supporting chimeras for each base-pair. *:
256 Regions with exceptionally low Shannon entropy; Vertical lines indicate long-
257 distance interactions. b, Viewpoint histograms showing binding positions of the
258 cyclization sequences along the ZIKV genome in control libraries; related to Fig. 2b.
259 Viewpoint regions are marked by dashed red lines. c, The non-circular ZIKV genome
260 conformation. Colour code as described in (a).
261
262 **Supplementary Fig. 5: Folding entropy.** a, Shannon entropy values calculated for
263 each nucleotide along the ZIKV genome. Entropy may range from 0 to 13.4 bits;
264 ZIKV coordinates are indicated by the position of genomic elements below. b,
265 Inverse correlation between the degree of experimental support of base paired
266 regions and their entropy. c, Shannon entropy values for a selected region of the
267 ZIKV genome. d, Number of supporting chimeric reads for each base-pair shown in
268 (c).
269
270 **Supplementary Fig. 6: Structures prediction.** Computationally predicted
271 structures for ~1,000 nucleotides regions along the ZIKV genome using randomly
272 selected subsets of experimentally probed folding-constraints. Each structure is
273 plotted as a dot according to its folding energy (dG) and experimentally supporting

evidence (chimera reads). Red dots indicating the structure with the lowest possible folding energy for each region.

Supplementary Fig. 7: Clustering of structures. a, Clustering of structures based on degree of similarity, related to Fig. 2h. Example of structures are shown for selected fragments with a colour code representing the number of non-redundant chimeric reads supporting each interaction. b, Representation of the % of *in vivo* probed interactions in use in an ensemble of 5 structures, an ensemble of 1,000 structures, or individual structures. Error bar represent standard deviation of 1,000 structures.

Supplementary Fig. 8: Co-clustering of non-shuffled and shuffled structures, related to Fig. 2h. a, Red scale indicates the number of chimeric reads supporting each non-shuffled structure. b, non-shuffled and shuffled structures, shown as white and black dots respectively.

Supplementary Fig. 9: *In vitro* affinity of Ago2-miR-21 to the ZIKV 5' CS. a, Electrophoretic mobility shift assay (EMSA) of 5'-labeled miR-21 (lanes 1, 2) upon addition of ZIKV 5' CS (lane 2); and 5'-labeled ZIKV 5' CS (lanes 3-5) upon addition of miR-21 (lane 4) or of Ago2-miR-21 complexes (lane 5). b-h, *In vitro* measured affinities of Ago2-miR-21 to a wildtype ZIKV 5' CS (b), mutated 5' CS (c-d), a fully seed matched mutated 5' CS (e), mutated 5' CS with no cHP stem-loop (f), extended 5' CS sequence containing the DCS stem-loop (g), and a short target perfectly complementary to the miR-21 seed (h).

Supplementary Fig. 10: miR-21 interaction with the ZIKV 5' CS. a, TaqMan PCR measurement of mature miR-21 expression level in wildtype cells and CRISPR/Cas9 miR-21 deletion-clones. Values are normalized to spike-in control. b, Intracellular ZIKV RNA in miR-21 knockout and wildtype cells. c, Expression level of control and miR-21 psiCHECK-2 reporters upon treatment with miR-21 or control inhibitors. miR-21 expression values denote for Renilla / Firefly luminescence signals. d, Intracellular ZIKV RNA in miR-21 inhibited and control cells. e-j, Intracellular levels of ZIKV envelope protein, measured by FACS. Representative experiment out of 3 is shown in (e-g); Gating strategy is shown in (h-j). NS1-AF488: Alexa Fluor antibody against ZIKV envelope protein. k, Replication of a ZIKV replicon carrying a wildtype 5' CS or a 5' CS - 3' CS double mutant, pre-treated with miR-21 or non-targeting inhibitors. wt: wildtype; KO1-2: CRISPR/Cas9 miR-21 deletion-clones 1 and 2; cont: control; Error bars represent standard deviation of 3 (a, b, d), 4 (c), and 6 (k) biological replicates; Two-sided Student's t-test p-value: ** <0.005; *** <0.0005; **** <0.00005; n.s.: non-significant.

Online methods

COMRADES. Each biological replicate was carried out independently on a different day, and included crosslinked sample, a control sample, and a non-crosslinked sample.

Psoralen crosslink. JEG-3 cells (~50 million cells per experiment) were inoculated with ZIKV isolate PE243, at MOI: 2 $TCID_{50}$/cell. 20 hours post inoculation cells were washed 3 times in HANKS solution and were incubated for 20 minutes with 0.4 mg/ml Psoralen-triethylene glycol azide (psoralen-TEG azide, Berry & Associates)

dissolved in PBS and diluted in OptiMEM I with no phenol-red (Gibco). Cells were irradiated on ice with 365 nM UV for 10 minutes using a CL-1000 crosslinker (UVP). Prolonged UVA irradiation should be avoided as it might decompose the azide moiety. Cells were lysed using RNeasy lysis buffer. Proteins were degraded by proteinase K (NEB) and RNA was purified using RNeasy midi kit (Qiagen).

Viral RNA enrichment. Total RNA was mixed with an array of 50 biotinylated DNA oligos, 20 nucleotides-long each (IDT) designed to capture the ZIKV genomic RNA and was maintained at 37 °C for 6 hours rotating in the following hybridization buffer: 500 mM NaCl, 0.7% SDS, 33 mM Tris-Cl pH 7, 0.7 mM EDTA, 10% Formamide. At the end of incubation Dynabeads MyOne Streptavidin C1(Invitrogen) were added and the RNA was incubated for additional 1 hour at 37 C. Beads were captured on a magnet and were washed 5 times with 2x SSC buffer containing 0.5% SDS. RNA was released from beads by degrading the DNA-probes with 0.1 units/µl Turbo DNase (Invitrogen) at 37 °C for 30 minutes. RNA was cleaned by RNA Clean & Concentrator (Zymo Research).

Crosslink pulldown. RNA was fragmented to an average size of 100 nucleotides using RNase III (Ambion) and was cleaned by RNA clean & concentrator (Zymo Research). Copper-free Click reaction was carried at 37 °C for 90 minutes in the presence of 150 µM Click-IT Biotin DIBO Alkyne (Life technologies) and 0.5 units/µl Superase-In (Invitrogen). Reaction was terminated by RNA Clean & Concentrator (Zymo Research). Biotinylated RNA was pulldown using Dynabeads MyOne Streptavidin C1(Invitrogen) at the following reaction conditions: 100 mM Tris-Cl pH 7.5, 10 mM EDTA, 1 M NaCl, 0.1% Tween-20, 0.5 unit/µl Superase-In. Beads were

captured on a magnet and were washed 5 times with 100 mM Tris-HCl pH 7.5, 10 mM EDTA, 3.5M NaCl, 0.1% Tween-20. RNA was eluted by adding 95% Formamide, 10 mM EDTA solution and incubating at 65oC for 5 minutes. To avoid enrichment of small RNA chimeric reads that cannot be double-aligned to the reference ZIKV genome / Human transcriptome, RNA was size fractionated on 10% TBE-Urea gel and fragments corresponding to a size of 100-200 nucleotides were eluted overnight at 4 °C in 10 mM Tris-HCl pH 7.5, 1 mM EDTA, 250 mM NaCl, 0.2% SDS. RNA was concentrated using RNA Clean & Concentrator (Zymo Research).

Proximity ligation and crosslink reversal. At this stage, the RNA sample was divided in two. One half was used for proximity ligation and then crosslink reversal (i.e. Crosslink sample), while in the other half, crosslink reversal was done before proximity ligation (i.e. Reverse-control). We included an additional control containing an equimolar concentration (albeit a non-similar composition) of non-psoralen treated, non-crosslinked enriched RNA (i.e., Non-crosslinked sample). Before proximity ligation, the RNA was heated to 85 °C for 2 minutes and was cooldown rapidly on ice. Proximity ligation was done under the following conditions: 1 unit/µl RNA ligase 1 (New England Biolabs), 1x RNA ligase buffer, 50 mM ATP, 1 unit/µl Superase-in (Invitrogen), final volume: 200 µl. Reaction was incubated for 16 hours at 16 °C and was terminated by cleaning with RNA Clean & Concentrator (Zymo Research). Crosslink reversal was done by irradiating the RNA on ice with 2.5KJ/m^2 UVC.

Sequencing library preparation. Library preparation was done as described in[8] with the following modifications: 6N unique molecular identifiers were added to the 5'

end of the 3' sequencing adapter; primers and adapters concentrations were lowered to match the low RNA input; Agencourt RNAClean XP beads (Beckman Coulter) were used for clean-up and size separation; pre-adenylated 5' and 3' adapters were used and all ligation reactions were carried without ATP to reduce ligation artefacts. All libraries and controls went through 13 PCR cycles using KAPA HiFi HotStart Ready Mix (KAPA Biosystems). PCR products were size-selected on a 1.8% agarose gel before loading on a HiSeq 1500 sequencer (Illumina).

Cell culture. JEG-3 placental trophoblasts (ATCC) and Hela cells (ATCC) were cultured in Minimum Essential Medium supplemented with 10% fetal bovine serum, 1 mM sodium pyruvate, GlutaMAX, non-essential amino acids and penicillin-streptomycin. Vero cells (Sigma-Aldrich) were cultured in Dulbecco's modified Eagle's medium supplemented with 10% fetal bovine serum, GlutaMAX and penicillin-streptomycin. All cell lines were cultured in a humidified CO_2 incubator at 37 °C and were regularly examined to exclude mycoplasma contamination.

Virus inoculation. ZIKV isolate PE243 was originated from Recife, Brazil in 2015. The virus was propagated in Vero cells and titer was determined by measuring the 50% Tissue Culture Infective Dose ($TCID_{50}$) in JEG-3 cells. For measurements of virus replication, JEG-3 or Hela cells were inoculated with ZIKV at MOI: 0.1 $TCID_{50}$/cell for three hours, after which cells were washed 3 times with PBS and supplemented with fresh growth medium. 24 hours post inoculation medium was removed, cells were washed 3 times with PBS and RNA was extracted using RNeasy kit (Qiagen). Virus copy number was determined using a TaqMan real-time PCR Assay (Primerdesign) and was normalized to GAPDH and ribosomal RNA. All

399 virus work was handled in a containment level 2 facility registered with the HSE
400 under COSHH.
401
402 **Replicon assay.** ZIKV wildtype and 5' CS - 3' CS double mutated replicons were
403 described previously[28]. Replicon RNA was synthesized *in vitro* using MEGAscript T7
404 Transcription Kit (Ambion). Replicon RNA was Capped using the ScriptCap m7G
405 Capping System (Cellscript) and transfected to Hela cells using the TransIT-mRNA
406 Transfection Kit (Mirus). Replicon levels were analyzed after 24-48 hours using a
407 microplate luminometer (Promega) and normalized to baseline luminescence values
408 measured at 6 hours post transfection.
409
410 **miR-21 knockout.** JEG-3 cells were transfected with CAS9-gRNA riboprotein
411 complexes using Lipofectamine RNAiMAX (Life technologies) according to the Alt-R
412 CRISPR-Cas9 user guide (IDT). miR-21 knockout clone1 was generated using the
413 following guide RNA: 5'-TCATGGCAACACCAGTCGATGGG-3' and contains a
414 homozygous deletion at the positions 59841310-59841326 on chromosome 17
415 (GRCh38/hg38 Assembly). miR-21 knockout clone2 was generated using a mixture
416 of two guide RNAs: 5'-ATGTCAGACAGCCCATCGACTGG-3', 5'-
417 CTACCATCGTGACATCTCCATGG-3', and contain a homozygous deletion at
418 positions 59841249-59841321 on chromosome 17. miR-21 knockout and control
419 clones were validated by sanger sequencing and by TaqMan Advanced miRNA
420 Assay targeting the mature miR-21 (Life Technologies).
421
422 **miR-21 inhibition.** Hela cells were transfected with inhibitors targeting human miR-
423 21 or non-targeting control A (Power inhibitors, Exiqon) at a final concentration of 25

nM using Lipofectamine RNAiMAX. 6 hours post transfection medium was replaced and cells were inoculated with ZIKV or re-transfected with ZIKV replicons as described above. miR-21 inhibition was validated using a psiCHECK-2 reporter (Promega) carrying a fully complementary miR-21 site at the 3' UTR of a Renilla luciferase reporter along with a Firefly reporter to normalise transfection efficiency. Luminescence was assessed using the Dual-reporter assay (Promega) and normalized to control psiCHECK-2 without the miR-21 binding site.

Gel-based Reverse Transcription Stalling (RTS) assay. RTS assay was performed as previously described[29] using a Cy5-labeled primer targeting the human 5.8S ribosomal RNA: 5'-Cy5-AAGCGACGCTCAGACAGG-3'.

Dot blot analysis. 50 ng crosslinked RNA, or the indicated amount of 50 nt-long biotinylated standards were spotted on to a Biodyne B Nylon Membrane (Life technologies) and dried by baking at 80 °C for 10 minutes. Biotinylated RNA was detected using the chemiluminescent nucleic acid detection module Kit (Life technologies) and visualized using ChemiDoc MP Imaging System (Biorad).

Purification of human Ago2 loaded with miR-21. Human Ago2 homogeneously loaded with miR-21 was prepared according to a published protocol[30]. Human Ago2 was expressed in Sf9 cells using the Bac-to-Bac baculovirus expression system (Thermo Fisher Scientific). Sf9 cells were lysed and human Ago2 was purified by Ni-NTA affinity chromatography using a His tag. Human Ago2 was loaded with synthetic 5'-phosphorylated miR-21 (IDT), and the His tag was removed using Tobacco etch virus protease. Human Ago2 loaded with miR-21 was captured using

449 an antisense oligonucleotide (IDT), eluted, and purified by size exclusion
450 chromatography on an ÄKTA FPLC (GE Healthcare Life Science). Protein
451 concentration was measured using absorption at 280 nM with extinction coefficients
452 obtained from the protparam tool (www.expasy.org) and from the ribotask
453 oligocalculator (www.ribotask.com).
454
455 **Target RNA labelling.** Synthetic RNA oligonucleotides (IDT) were radiolabelled at
456 the 5'-end using gamma 32P ATP (Perkin Elmer) and T4 polynucleotide kinase
457 (NEB), and purified by denaturing polyacrylamide gel and ethanol precipitation. RNA
458 concentration was determined from absorption at 260 nM using extinction
459 coefficients calculated with the ribotask oligocalculator (www.ribotask.com).
460
461 **Electrophoretic mobility shift assay (EMSA).** Binding reactions were prepared in
462 reaction buffer (28 mM Tris pH 8.0, 20 mM KCl, 80 mM KOAc, 1.6 mM Mg(OAc)2,
463 0.5 mM TCEP, 0.004% NP-40, 0.01 g/l baker's yeast tRNA) with a final volume of 20
464 µl, and a final concentration of the labeled RNAs of 10 nM and of the non-labeled
465 RNA or Ago2-miR-21 of 100 nM. Reactions were incubated for 10 minutes at room
466 temperature and analyzed on a 15% acrylamide native gel in 0.5x TBE.
467
468 **Kd measurements.** Binding experiments were conducted according to the protocol
469 published in[30]. Ago2-miR-21 (0-200 nM) was incubated with 0.1 nM radiolabaled
470 target in reaction buffer (28 mM Tris pH 8.0, 20 mM KCl, 80 mM KOAc, 1.6 mM
471 Mg(OAc)2, 0.5 mM TCEP, 0.004% NP-40) with a total volume of 25 µl for 45 minutes
472 at room temperature. Filter-binding was performed using a dot-blot apparatus (GE
473 Healthcare Life Sciences) with Protran nitrocellulose membrane (Amersham, GE

Healthcare Life Sciences) and Hybond N+ nylon membrane (Amersham, GE Healthcare Life Sciences). Samples were applied with vacuum and washed with 50 µl wash buffer (30 mM Tris pH 8.0, 100 mM KOAc, 2 mM Mg(OAc)2, 0.5 mM TCEP). After air drying, the membrane strips were used to expose phosphor screens (GE Healthcare Life Sciences) for visualization. Screens were imaged on a Typhoon phosphorimager (GE Healthcare Life Science) and signals were quantified with ImageQuant (GE Healthcare Life Sciences). Dissociation constants were calculated by fitting the data to a single site binding equation:

$$F = \frac{Bmax[Ago2]}{[Ago2] + KD}$$

F = fraction target RNA bound, Bmax = maximal number of binding sites, [Ago2] = total concentration of the Ago2-miR21 complex, and KD = calculated dissociation constant, using Prism (GraphPad Software). For weakly binding RNAs Bmax was constrained to ≤1.

Processing and visualization of sequencing data. Sequencing data were pre-processed to combine FASTQ files of two sequencing lanes (cat) and to remove adapters (cutadapt). Paired end reads were merged by paired-end read merger (pear). UMIs were collapsed by collapse.py (T.D. Domenico, https://github.com/tdido). Chimeric reads were called and annotated with the hyb package[31], using the command:

```
hyb analyse in=data.fasta db=hOH7_and_Zika format=comp eval=0.001
```

Hyb uses bowtie2[32] in local mapping mode to map reads to a transcriptome database and to identify chimeras, and it annotates the chimeras with RNA base-

499 pairing information generated by hybrid-min[33]. The transcriptome database used by
500 hyb, “hOH7_and_Zika”, consists of human spliced mRNAs and noncoding RNAs
501 described in[14], and the genome sequence of the Zika virus (Zika virus isolate
502 ZIKV/H.sapiens/Brazil/PE243/2015, complete genome). To evaluate the folding
503 energy of chimeric reads, we used hybrid-min[33] with default settings. We then
504 randomly reassigned (shuffled) pairs of fragments found in chimeric reads, and
505 repeated the folding energy analysis. The folding energies of experimentally
506 identified and shuffled chimeras were compared by Wilcoxon test.
507
508 Virus interaction heatmaps were plotted using Java Treeview[34], such that color
509 intensity represents the coverage of chimeric reads at every pair of positions. The
510 first read of each pair is plotted along the X axis, and the second read along the Y
511 axis. As a result, chimeras found in the 5'-3' orientation are shown above the
512 diagonal, and chimeras in the 3'-5' orientation are below the diagonal. Viewpoint
513 histograms were plotted with gnuplot, and arc plots were plotted with R-chie[35].
514
515 For every pair of positions *(i, j)* along the virus genome we calculated the
516 COMRADES score, C_{ij}: the number of chimeric reads that, when analyzed with the
517 program hybrid-min with default settings, indicated base-pairing between positions *i*
518 and *j*. We used COMRADES scores to calculate per-base Shannon entropy for each
519 nucleotide position along the virus. Shannon entropy of position *i* is defined as:

520 $$Entropy_{\,i} = -\sum_{j=1}^{n} P(C_{ij}) log_2 P(C_{ij}),$$

521 where n is the length of the genome (10,807 nt); and P(Cij) is

$$P(C_{ij}) = C_{ij} \Big/ \sum_{k=1}^{n} C_{ik} .$$

High entropy indicates flexible positions that may form multiple alternative base-pairs, whereas low entropy indicates positions that always pair with the same nucleotide partner. We visualized RNA structures using VARNA[36], where the colour scale represents the COMRADES score for each base pair.

RNA structure prediction. For RNA structure predictions, we collected all potential base pairs with a non-zero C_{ij} value, assembled sets of adjacent base pairs into uninterrupted stem structures, and calculated the base-pairing score of each stem as the sum of C_{ij} values of individual base pairs. We then ranked these stem elements by their scores. In a preliminary analysis, we folded the 10,807 nt virus genome in a set of 50 overlapping 1,000 nt fragments, using the hybrid-ss-min program[33]. Each fragment was folded using a set of 250 top-ranked *in vivo* probed stem elements as folding constraints. Based on this preliminary analysis, we identified high-scoring stem-loop structures that were reproducibly predicted across multiple fragments, and we defined new fragment boundaries to prevent the disruption of these reproducible structural elements. As a result, we obtained fragment sizes that vary in size, but are approximately 1,000 nt long each.

We then performed full folding analysis using the following fragment boundaries: 5'UTR: 1-107, F1: 108-1275, F2: 1276-2287, F3: 2288-3323, F4: 3324-4521, F5: 4522-5551, F6: 5552-6810, F7: 6811-7757, F8: 7758-8755, F9: 8756-9543, F10: 9544-10379, 3'UTR: 10380-10807.

For each fragment, we assembled a set of folding constraints that represented the 75 top-scoring stem elements within that fragment, we randomly shuffled this set of constraints 1,000 times, and we used the shuffled constraints for folding prediction by hybrid-ss-min. The resulting individual structures typically incorporate 25%-40% of these constraints. We recorded the folding energy of each structure, as predicted by hybrid-ss-min, and we used the sum of C_{ij} values to calculate an overall score for each structure. To assemble the top-scoring full-genome structure shown in Supplementary Fig. 4a, we assembled the top-scoring structures for each coding sequence fragment (F1-F10), and the previously proposed structures of the 5' and 3' UTRs. An additional analysis of folding within and between the 5' and 3' UTRs is shown in Fig. 2a-c and Supplementary Fig. 4c.

We also repeated the folding analysis with shuffled sets of 50-250 top-scoring constraints per fragment. This yielded similar results, but we found that either reducing or increasing the numbers of constraints tended to reduce the number of high-scoring structures.

To explore the sets of alternative structures, we computed pairwise distances between structures as the number of positions with discordant base-pairing. This resulted in a 1,000 x 1,000 matrix of distances, which we then represented on a two-dimensional surface using multidimensional scaling (using the R function cmdscale). Multidimensional scaling, also known as Principal Coordinate Analysis[37], maps multidimensional objects (in this case, RNA structures) to a set of points on a plane, such that the distances between RNA structures are well-approximated by Euclidean distances between points, by minimization of a stress function:

570 $$Stress_D(x_1,\ldots,x_n) = \left(\sum_{i\neq j=1..n}(D_{i,j} - \|x_i - x_j\|)^2\right)^{1/2}$$

571

572 The RNA structure prediction pipeline can be downloaded from:

573 https://github.com/gkudla/comrades

574

575 **Validation of ZIKV miR-21 interaction with an independent analysis pipeline.**

576 **Alignment.** The first read of each pair was processed using UMI-tools[38] to extract

577 the 6 nucleotide unique molecular identifier (UMI) at the start of the read. Processed

578 reads were aligned using the STAR aligner[39], reporting all reads in their original

579 order (--outSAMtype BAM Unsorted --outSAMunmapped Within); only reporting

580 unique alignments (--outFilterMultimapNmax 1); and reporting alignments to

581 individual segments of chimeric reads (--chimOutType WithinBAM --

582 chimSegmentMin 20 --chimScoreJunctionNonGTAG 0 --

583 chimMainSegmentMultNmax 1). The reference consisted of the hg38 build of the

584 human genome, combined with the genome sequence of the PE243 strain of the

585 Zika virus. Each read of the pair was aligned separately to avoid preferencing

586 alignment to the same genomic locus.

587 For each library, the pair of BAM files were collated and pair information was fixed

588 using samtools[40]. PCR duplicates were removed on the basis of their UMIs, using

589 UMI-tools in paired mode.

590

591 **Detecting significant interactions.** We considered the "interaction space" between

592 the human and Zika genomes, consisting of pairs of 1 kbp bins (one on each

593 genome). For each replicate library in each condition (crosslinked and reversed

594 control), we counted the number of read pairs with one read in each bin using

diffHic[41]. This yielded a count matrix that was normalized using the trimmed mean-of-M-values method[42] to correct for composition biases, under the assumption that most read pairs mapping across the Zika and human genomes was caused by non-specific ligation. We then applied the quasi-likelihood framework in edgeR[43] to test for significant differences between the read pair counts for crosslinked and the reverse control. This was performed using an additive design matrix that blocked on the batch to reflect the paired-sample design of the experiment.

Robust empirical Bayes shrinkage[44] was also used to stabilise the dispersion estimates in the presence of limited replication. Bin pairs were aggregated into clusters based on whether they overlapped the same human gene

Test statistics were combined for each gene-Zika interaction using Simes' method[45] prior to applying the Benjamini-Hochberg method. Interactions that were significantly enriched in COMRADES over the reversed control were defined at a false discovery rate threshold of 5%.

References

1. Wan, Y. *et al.* Landscape and variation of RNA secondary structure across the human transcriptome. *Nature* **505,** 706–709 (2014).
2. Ding, Y. *et al.* In vivo genome-wide profiling of RNA secondary structure reveals novel regulatory features. *Nature* **505,** 696–700 (2013).
3. Rouskin, S., Zubradt, M., Washietl, S., Kellis, M. & Weissman, J. S. Genome-wide probing of RNA structure reveals active unfolding of mRNA structures in vivo. *Nature* **505,** 701–705 (2014).
4. Spitale, R. C. *et al.* Structural imprints in vivo decode RNA regulatory mechanisms. *Nature* **519,** 486–490 (2015).

5. Mustoe, A. M. *et al.* Pervasive Regulatory Functions of mRNA Structure Revealed by High-Resolution SHAPE Probing. *Cell* **173,** 181–195.e18 (2018).
6. Watts, J. M. *et al.* Architecture and secondary structure of an entire HIV-1 RNA genome. *Nature* **460,** 711–716 (2009).
7. Pirakitikulr, N., Kohlway, A., Lindenbach, B. D. & Pyle, A. M. The Coding Region of the HCV Genome Contains a Network of Regulatory RNA Structures. *Mol. Cell* **62,** 111–120 (2016).
8. Kwok, C. K., Marsico, G., Sahakyan, A. B., Chambers, V. S. & Balasubramanian, S. rG4-seq reveals widespread formation of G-quadruplex structures in the human transcriptome. *Nat. Methods* **13,** 841–844 (2016).
9. Li, H. & Aviran, S. Publisher Correction: Statistical modeling of RNA structure profiling experiments enables parsimonious reconstruction of structure landscapes. *Nat. Commun.* **9,** 1110 (2018).
10. Lu, Z. *et al.* RNA Duplex Map in Living Cells Reveals Higher-Order Transcriptome Structure. *Cell* **165,** 1267–1279 (2016).
11. Aw, J. G. A. *et al.* In Vivo Mapping of Eukaryotic RNA Interactomes Reveals Principles of Higher-Order Organization and Regulation. *Mol. Cell* **62,** 603–617 (2016).
12. Sharma, E., Sterne-Weiler, T., O'Hanlon, D. & Blencowe, B. J. Global Mapping of Human RNA-RNA Interactions. *Mol. Cell* **62,** 618–626 (2016).
13. Kudla, G., Granneman, S., Hahn, D., Beggs, J. D. & Tollervey, D. Cross-linking, ligation, and sequencing of hybrids reveals RNA–RNA interactions in yeast. *Proc. Natl. Acad. Sci. U. S. A.* **108,** 10010–10015 (2011).

643 14. Helwak, A., Kudla, G., Dudnakova, T. & Tollervey, D. Mapping the human
644 miRNA interactome by CLASH reveals frequent noncanonical binding. *Cell*
645 **153,** 654–665 (2013).
646 15. Ramani, V., Qiu, R. & Shendure, J. High-throughput determination of RNA
647 structure by proximity ligation. *Nat. Biotechnol.* **33,** 980–984 (2015).
648 16. Sugimoto, Y. *et al.* hiCLIP reveals the in vivo atlas of mRNA secondary
649 structures recognized by Staufen 1. *Nature* **519,** 491–494 (2015).
650 17. Liu, Z.-Y. *et al.* Viral RNA switch mediates the dynamic control of flavivirus
651 replicase recruitment by genome cyclization. *Elife* **5,** (2016).
652 18. Hahn, C. S. *et al.* Conserved elements in the 3’ untranslated region of
653 flavivirus RNAs and potential cyclization sequences. *J. Mol. Biol.* **198,** 33–41
654 (1987).
655 19. Alvarez, D. E., Lodeiro, M. F., Ludueña, S. J., Pietrasanta, L. I. & Gamarnik,
656 A. V. Long-range RNA-RNA interactions circularize the dengue virus genome.
657 *J. Virol.* **79,** 6631–6643 (2005).
658 20. Friebe, P. & Harris, E. Interplay of RNA elements in the dengue virus 5’ and 3'
659 ends required for viral RNA replication. *J. Virol.* **84,** 6103–6118 (2010).
660 21. Filomatori, C. V. *et al.* A 5’ RNA element promotes dengue virus RNA
661 synthesis on a circular genome. *Genes Dev.* **20,** 2238–2249 (2006).
662 22. Manzano, M. *et al.* Identification of Cis-Acting Elements in the 3′-Untranslated
663 Region of the Dengue Virus Type 2 RNA That Modulate Translation and
664 Replication. *J. Biol. Chem.* **286,** 22521–22534 (2011).
665 23. Liu, Z.-Y. *et al.* Novel cis-acting element within the capsid-coding region
666 enhances flavivirus viral-RNA replication by regulating genome cyclization. *J.*
667 *Virol.* **87,** 6804–6818 (2013).

668 24. Akiyama, B. M. *et al.* Zika virus produces noncoding RNAs using a multi-
669 pseudoknot structure that confounds a cellular exonuclease. *Science* **354,**
670 1148–1152 (2016).
671 25. Ding, Y. *et al.* In vivo genome-wide profiling of RNA secondary structure
672 reveals novel regulatory features. *Nature* **505,** 696–700 (2014).
673 26. Guo, J. U. & Bartel, D. P. RNA G-quadruplexes are globally unfolded in
674 eukaryotic cells and depleted in bacteria. *Science* **353,** (2016).
675 27. Guo, Y. E. & Steitz, J. A. Virus meets host microRNA: the destroyer, the
676 booster, the hijacker. *Mol. Cell. Biol.* **34,** 3780–3787 (2014).
677

678 **Supplementary references**

679 28. Liu, Z.-Y. *et al.* Characterization of cis-acting RNA elements of Zika virus by
680 using a self-splicing ribozyme-dependent infectious clone. *J. Virol.* **91,**
681 e00484–17 (2017).
682 29. Kwok, C. K., Ding, Y., Tang, Y., Assmann, S. M. & Bevilacqua, P. C.
683 Determination of in vivo RNA structure in low-abundance transcripts. *Nat.*
684 *Commun.* **4,** 2971 (2013).
685 30. Schirle, N. T., Sheu-Gruttadauria, J. & MacRae, I. J. Structural basis for
686 microRNA targeting. *Science* **346,** 608–613 (2014).
687 31. Travis, A. J., Moody, J., Helwak, A., Tollervey, D. & Kudla, G. Hyb: a
688 bioinformatics pipeline for the analysis of CLASH (crosslinking, ligation and
689 sequencing of hybrids) data. *Methods* **65,** 263–273 (2014).
690 32. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2.
691 *Nat. Methods* **9,** 357–359 (2012).

692 33. Markham, N. R. & Zuker, M. UNAFold. in *Bioinformatics: Structure, Function*
693 *and Applications* (ed. Keith, J. M.) 3–31 (Humana Press, 2008).
694 34. Saldanha, A. J. Java Treeview—extensible visualization of microarray data.
695 *Bioinformatics* **20,** 3246–3248 (2004).
696 35. Lai, D., Proctor, J. R., Zhu, J. Y. A. & Meyer, I. M. R-CHIE: a web server and
697 R package for visualizing RNA secondary structures. *Nucleic Acids Res.* **40,**
698 e95 (2012).
699 36. Darty, K., Denise, A. & Ponty, Y. VARNA: Interactive drawing and editing of
700 the RNA secondary structure. *Bioinformatics* **25,** 1974–1975 (2009).
701 37. Gower, J. C. Some distance properties of latent root and vector methods used
702 in multivariate analysis. *Biometrika* **53,** 325–338 (1966).
703 38. Smith, T., Heger, A. & Sudbery, I. UMI-tools: modeling sequencing errors in
704 Unique Molecular Identifiers to improve quantification accuracy. *Genome Res.*
705 **27,** 491–499 (2017).
706 39. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29,**
707 15–21 (2013).
708 40. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools.
709 *Bioinformatics* **25,** 2078–2079 (2009).
710 41. Lun, A. T. L. & Smyth, G. K. diffHic: a Bioconductor package to detect
711 differential genomic interactions in Hi-C data. *BMC Bioinformatics* **16,** 258
712 (2015).
713 42. Robinson, M. D. & Oshlack, A. A scaling normalization method for differential
714 expression analysis of RNA-seq data. *Genome Biol.* **11,** R25 (2010).

43. Lun, A. T. L., Chen, Y. & Smyth, G. K. It's DE-licious: A Recipe for Differential Expression Analyses of RNA-seq Experiments Using Quasi-Likelihood Methods in edgeR. *Methods Mol. Biol.* **1418,** 391–416 (2016).

44. Phipson, B., Lee, S., Majewski, I. J., Alexander, W. S. & Smyth, G. K. Robust hyperparameter estimation protects against hypervariable genes and improves power to detect differential expression. *Ann. Appl. Stat.* **10,** 946–963 (2016).

45. Lun, A. T. L. & Smyth, G. K. De novo detection of differentially bound regions for ChIP-seq data using peaks and windows: controlling error rates correctly. *Nucleic Acids Res.* **42,** e95 (2014).

Acknowledgements The authors thank Alain Kohl (Centre for Virus Research, University of Glasgow) and Lindomar J. Pena and Rafael Oliveira de Freitas França, Fiocruz Recife, Pernambuco, Brazil, for providing the PE243 ZIKV RNA used to generate the virus stock. We thank Yaron Galanty and Francisco M. Martínez for assisting with CRISPR knockouts; Tomás D. Domenico and Wayo Matsushima for collapsing U.M.Is; Guido Sanguinetti for suggesting the RNA folding strategy; Amit Zeisel for designing ZIKV probes; Sylviane Moss for assisting with risk assessments; Michael S. Diamond, Trevor Sweeney, Andrew Firth, David Jordan, and members of E.A.M. group for their comments and Claudia Flandoli for illustrations. This work was supported by Cancer Research UK (C13474/A18583, C6946/A14492) and the Wellcome Trust (104640/Z/14/Z, 092096/Z/10/Z) to E.M. O.Z. was supported by the Human Frontier Science Program (HFSP, LT000558/2015), the European Molecular Biology Organization (EMBO, ALTF1622-2014), and the Blavatnik Family Foundation postdoctoral fellowship. G.K. and M.G. were supported by Wellcome Trust grant 207507 and UK Medical Research Council. A.T.L.L. and J.C.M. were

740 supported by core funding from Cancer Research UK (award no. 17197 to JCM).
741 J.C.M was also supported by core funding from EMBL. I.G. and L.W.M. were
742 supported by the Wellcome Trust Senior Fellowship in Basic Biomedical Science to
743 I.G. (207498/Z/17/Z). I.J.M., L.F.G. and J.S.-G. were supported by grants
744 R01GM104475 and R01GM115649 from NIGMS. C.K.K was supported by City
745 University of Hong Kong Projects 9610363 and 7200520, Croucher Foundation
746 Project 9500030 and Hong Kong RGC Projects 9048103 and 9054020. C.-F.Q. was
747 supported by the NSFC Excellent Young Scientist Fund 81522025 and the Newton
748 Advanced Fellowship from the Academy of Medical Sciences, UK.

749

750 **Author contributions** O.Z., A.T.L.L., J.C.M. and E.A.M. designed the study; O.Z.
751 developed COMRADES and performed *in vivo* experiments with assistance from
752 L.W.M., I.G. and C.K.K.; L.F.G., J.S.-G. and I.J.M. preformed *in vitro* binding
753 experiments; Z.-Y.L. and C.-F.Q. provided ZIKV clones and replicons; M.M.G.,
754 A.T.L.L., O.Z., E.A.M., G.K. and J.C.M. performed data analysis; O.Z., G.K. and
755 E.A.M. wrote the paper with input from all authors.

756

757 **Author information** All sequencing data sets have been deposited in ArrayExpress
758 under accession number: E-MTAB-6427. Reprints and permissions information is
759 available at www.nature.com/reprints. The authors declare no competing financial
760 interests. Readers are welcome to comment on the online version of the paper.
761 Correspondence and requests for materials should be addressed to E.A.M
762 (eric.miska@gurdon.cam.ac.uk) or G.K (gkudla@gmail.com) or O.Z.
763 (omer.ziv@gurdon.cam.ac.uk).